# VU Research Portal

## The International Index of Erectile Function (IIEF)-A Systematic Review of Measurement Properties

Neijenhuijs, Koen I.; Holtmaat, Karen; Aaronson, Neil K.; Holzner, Bernhard; Terwee, Caroline B.; Cuijpers, Pim; Verdonck-de Leeuw, Irma M.

**Link to publication in VU Research Portal**

# The International Index of Erectile Function (IIEF)—A Systematic Review of Measurement Properties

Koen I. Neijenhuijs, MSc,[1] Karen Holtmaat, MSc,[1] Neil K. Aaronson, Prof,[2] Bernhard Holzner, Prof,[3] Caroline B. Terwee, PhD,[4] Pim Cuijpers, Prof,[1] and Irma M. Verdonck-de Leeuw, Prof[1,5]

## ABSTRACT

**Introduction:** The International Index of Erectile Function (IIEF) is a patient-reported outcome measure to evaluate erectile dysfunction and other sexual problems in men.

**Aim:** To perform a systematic review of the measurement properties of the 15-item patient-reported outcome measure (IIEF-15) and the shortened 5-item version (IIEF-5).

**Methods:** A systematic search of scientific literature up to April 2018 was performed. Data were extracted and analyzed according to COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines for structural validity, internal consistency, reliability, measurement error, hypothesis testing for construct validity, and responsiveness. Evidence of measurement properties was categorized into sufficient, insufficient, inconsistent, or indeterminate, and quality of evidence as very high, high, moderate, or low.

**Results:** 40 studies were included. The evidence for criterion validity (of the Erectile Function subscale), and responsiveness of the IIEF-15 was sufficient (high quality), but inconsistent (moderate quality) for structural validity, internal consistency, construct validity, and test-retest reliability. Evidence for structural validity, test-retest reliability, construct validity, and criterion validity of the IIEF-5 was sufficient (moderate quality) but indeterminate for internal consistency, measurement error, and responsiveness.

**Clinical Implications:** Lack of evidence for and evidence not supporting some of the measurement properties of the IIEF-15 and IIEF-5 shows the importance of further research on the validity of these questionnaires in clinical research and clinical practice.

**Strengths & Limitations:** A strength of the current review is the use of predefined guidelines (COSMIN). A limitation of this review is the use of a precise rather than a sensitive search filter regarding measurement properties to identify studies to be included.

**Conclusion:** The IIEF requires more research on structural validity (IIEF-15), internal consistency (IIEF-15 and IIEF-5), construct validity (IIEF-15), measurement error (IIEF-15 and IIEF-5), and responsiveness (IIEF-5). The most pressing matter for future research is determining the unidimensionality of the IIEF-5 and the exact factor structure of the IIEF-15. **Neijenhuijs KI, Holtmaat K, Aaronson NK, et al. The International Index of Erectile Function (IIEF)—A Systematic Review of Measurement Properties. J Sex Med 2019;16:1078−1091.**

Copyright © 2019, International Society for Sexual Medicine. Published by Elsevier Inc. All rights reserved.

**Key Words:** International Index of Erectile Function; Validity; Reliability; COSMIN; Measurement Properties

# INTRODUCTION

The International Index of Erectile Function (IIEF) is a widely used patient-reported outcome measure (PROM) to evaluate sexual problems in men.[1] The IIEF is a 15-item PROM (IIEF-15) including 5 domains: erectile function (6 items), orgasmic function (2 items), sexual desire (2 items), intercourse satisfaction (3 items), and overall satisfaction (2 items). Initial research revealed that the IIEF-15 had acceptable internal consistency ($\alpha > 0.70$) and test-retest reliability ($r > 0.70$), except for the orgasmic function scale.[1] Construct validity was good, and the IIEF-15 could detect changes before and after treatment.[1] A shortened 5-item version was developed to evaluate sexual problems in men by selecting the items that best discriminated between men with and without erectile dysfunction (ED) and adhered to the National Institutes of Health's definition of ED. The result was a 5-item version consisting of 4 items from the erectile function, and 1 item from the sexual intercourse satisfaction subscales. The IIEF-5 was able to discriminate clearly between patients with ED and those without.[2]

Information regarding validity and reliability is of importance for clinical research and practice. To be able to interpret the IIEF-15 and IIEF-5, we need to be certain that the subscales measure what they intend to measure, that they do so consistently, and (particularly for practice) what cutoff scores can be used to screen patients for ED. A review published in 2002 concluded that the IIEF was translated in 32 languages and adopted as a primary endpoint in >50 clinical trials worldwide.[3] The authors reported that the IIEF-15 met the standard psychometric criteria for reliability and validity, had a high degree of sensitivity and specificity, and correlated well with other measures of treatment outcome. It also demonstrated good responsiveness.[3]

However, since then, many more studies have been published investigating the psychometric properties of the IIEF-15 and IIEF-5. Given the high frequency of use in both clinical practice and research, an update of the evidence on the psychometric properties of the IIEF-15 and IIEF-5 is warranted to investigate whether the initial results[1-3] have been replicated in independent international and more recent validation studies. Therefore, the aim of this study was to perform a systematic review of the measurement properties of the IIEF-15 and IIEF-5.

In this review, we followed the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology.[4] This methodology is based on taxonomy and definitions of measurement properties for PROMs,[5] including content validity, structural validity, internal consistency, cross-cultural validity, reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness. We hypothesized that there would be evidence supporting sufficient psychometric values IIEF-15 and IIEF-5.

# METHODS

## Literature Search Strategy

The literature search was part of a larger systematic review (Prospero ID 42017057237), which investigated the measurement properties of 39 PROMs (including the IIEF-15 and IIEF-5) assessing the quality of life of cancer survivors included in an eHealth application called "Oncokompas".[6-10] The databases Embase, Medline, and Web of Science were searched using the search terms of the PROM's name and acronyms, combined with a precise filter for measurement properties.[10] The search was performed in January 2017. Appendix A contains the full search terms with regard to all 39 PROMs. Appendix B contains the search terms relating specifically to the IIEF. References were extracted from systematic reviews found in an earlier search of the larger systematic review, and added to the search results. A search update was performed in April 2018. Due to the limitation of the sensitivity of the precise filter (93% sensitive),[10] a manual search using rudimentary search filters was performed in Google Scholar and PubMed to check for any prominent records missed in the search update.

## Inclusion and Exclusion Criteria

Studies were included that reported original data on ≥1 of the following measurement properties of the IIEF as defined by the COSMIN taxonomy[5,11,12]: structural validity (whether the hypothesized measurement model is confirmed), internal consistency (the degree of interrelatedness among the items of the measure), reliability (the proportion of total variance between multiple measurements, which is due to "true" differences between measurements), measurement error (a measure of systematic and random error in change scores), criterion validity (whether the measure is an adequate reflection of a gold standard; in the case of the IIEF this is most often a diagnosis of ED), cross-cultural validity (whether the test can be interpreted similarly in different cultures), responsiveness (whether the measure is capable of measuring change over time in the construct to be measured), and hypothesis testing for construct validity (whether the test measures the construct it proposes to measure), which consists of known-groups comparison (a comparison between groups known to have differences on the construct), convergent validity (correlations with other measures that should be related), and divergent validity (correlations with other measures that should be unrelated). Although of importance for establishing validity, content validity was not investigated because it was beyond the scope of the current review. Validation studies focused on other PROMs, and non-validation studies that used the IIEF that also reported evidence on the measurement properties of the IIEF were included.

Studies that were only available as abstracts or conference proceedings were excluded, as well as non-English publications. Titles and abstracts, and the selected full-texts were screened by 2 independent reviewers (K.N. & M.V./K.H.). Disagreements were discussed until consensus was reached.

## Data Extraction

Data on each of the measurement properties was extracted by two independent researchers (K.N. & A.vdH./H.M./E.V./K.H.). Relevant data included the type of measurement property, its result, and information on methodology. Disagreements were discussed until consensus was reached.

## Data Analysis

Data analysis was performed in 3 consecutive steps. First, the methodologic quality of the included studies was rated using the 4-point scoring system of the COSMIN checklist.[13] Methodologic aspects regarding design requirements and preferred statistical methods specific to each measurement property under consideration, were rated as either "inadequate," "doubtful," "adequate," or "very good." The methodologic quality was summarized per measurement property per study as the lowest score received on any of the methodologic aspects. Appendix C contains the final study quality ratings.

Second, each measurement property in each individual study was rated as sufficient, insufficient or indeterminate, following the COSMIN guidelines for systematic reviews of PROMs.[4] These ratings were qualitatively summarized to determine the overall rating of the measurement property for the IIEF. If all studies indicated a "sufficient," "insufficient," or "indeterminate" rating for a specific measurement property, the overall rating of this measurement property was rated accordingly. If there were inconsistencies between studies, explanations were explored (eg, differences in methodologic quality, differences in population, etc). If explanations were found, they were discussed until consensus was reached regarding the overall rating of the measurement property. If no explanations were found, the overall rating would be inconsistent.

Third, the overall rating of evidence per measurement property was supplemented by a level of quality of the evidence, using a modified Grading of Recommendations Assessment, Development and Evaluation approach from the COSMIN methodology.[4] This approach takes into account (i) study quality, (ii) directness of evidence, (iii) inconsistency of results, and (iv) precision of evidence (number of studies and sample size). The overall quality of evidence was rated as high, moderate, low, or very low. Measurement properties that were rated as indeterminate in the previous step did not receive a rating, as there was no evidence to rate.

All ratings (methodologic quality, measurement property rating, and Grading of Recommendations Assessment, Development and Evaluation rating) were rated by 2 independent researchers (K.N. & K.H.). Discrepancies in ratings were discussed until consensus was reached.

## RESULTS

### Search Results

The initial search identified 1,401 non-duplicate abstracts of which 568 were relevant to the IIEF (Supplementary Figure 1). A total of 526 abstracts and 17 full texts were excluded because

they did not provide unique information on a measurement property. The search update up to April 2018 identified 342 more non-duplicate abstracts. A total of 317 abstracts and 17 full texts were excluded because they did not provide unique information on a measurement property of the IIEF. 10 references were found through manual means, of which 5 were excluded during abstract screening because they did not provide unique information on a measurement property of the IIEF.

In total, we included 40 articles: 31 on the IIEF-15,[1,14–43] 7 on the IIEF-5,[2,44–49] and 2 on both the IIEF-15 and IIEF-5.[50,51] An overview of study characteristics is provided in Table 1. Studies reported sample sizes ranging from 40 to 1,764, and 12 different countries were reported: Turkey (Turkish), Spain (Spanish), Taiwan (Taiwanese Mandarin/Hokkien), Germany (German), Iran (Persian), Italy (Italian), Malaysia (Malay), Portugal (Portuguese), China (Chinese), Canada (French), Pakistan (Urdu), and the Netherlands (Dutch). Other included studies likely have been conducted in other countries, but the nationality of participants was not always clearly specified. The combined body of the 33 studies on the IIEF-15 and the 9 studies on the IIEF-5 reported on all measurement properties, except cross-cultural validity.

### Structural Validity

8 studies reported on structural validity of the IIEF-15,[1,17,22,26,28,36,43,51] of which 1 study[36] reported 2 types of analyses (Table 2). Methodologic quality was rated as "very good",[17,28] "adequate",[1,22,43,51] or "doubtful".[26,36] 1 "doubtful" score was due to an insufficient sample size ("other flaws" in COSMIN methodologic quality),[26] whereas the other was because of very unequal subgroup sizes ("other flaws" in COSMIN methodologic quality).[36]

3 studies of "very good"[17,28] and "doubtful"[36] quality reported confirmatory factor analyses (CFAs). The evidence on structural validity was rated as sufficient in 2 studies, because a good fit was found for a 5-factor structure.[28,36] The evidence was rated as insufficient for the third study, because the fit for the 5-factor structure was below acceptable levels (Comparative Fit Index [CFI] < 0.95).[17] The evidence was rated as indeterminate for 6 studies of the IIEF-15, of "adequate"[1,22,43,51] and "doubtful"[26,36] quality, because they reported principal component analyses (PCAs) without fit measures. Notably, 2 of these studies reproduced the hypothesized 5 components, 2 studies found 4 components, and 2 studies found 2 components.

1 study reported on structural validity of the IIEF-5[46] (Table 2). Methodologic quality was rated as "very good." Evidence on structural validity was rated as sufficient, because a good fit of a Rasch model was reported.

### Internal Consistency

15 studies reported on internal consistency of the IIEF-15[1,15,16,22,26,28,31,34–36,38,41,43,50,51] (Supplementary Table 1). Methodologic quality of these studies was rated as "very

**Table 1.** Characteristics of included studies

| Reference | Population | Sample size | Main aim of study |
|---|---|---|---|
| **IIEF-15** | | | |
| Althof et al[14] | Patients with ED with somewhat low self-esteem | 282 | Investigate the impact of sildenafil treatment on psychosocial functioning and well-being in men with ED from 4 countries |
| Bayraktar et al[15] | Patients with ED | 225 | Assess the reliability of the physician-assisted IIEF-15 (Turkish version) in patients with ED |
| Bayraktar et al[16] | Patients with ED | 458 | To analyze the impact of assistance on the comprehensibility and reliability of the Turkish version of the IIEF-15 questionnaire |
| Bushmakin et al[17] | Patients with ED enrolled in a RCT on sildenafil | 500 | Testing structural validity of IIEF-15 |
| Cappelleri et al[18] | 111 ED patients in RCT on sildenafil; 109 control patients; 37 ED patients; and 21 age-matched controls | 278 | Development and validation of IIEF-15 |
| Cappalleri et al[19] | Patients with ED enrolled in a RCT on sildenafil | 247 | Examine the relationship between patients' self-assessment of EF and the EF domain of the IIEF with respect to ED severity |
| Cappalleri et al[20] | Patients with ED enrolled in a RCT on sildenafil | 209 | Mapping the relationship among 4 categories of the EHS and the IIEF-EF, QEQ, SEX-Q, and SEAR |
| O'Leary et al[21] | Patients with ED enrolled in a RCT on sildenafil with somewhat low self-esteem | 244 | Assess the change in confidence, relationship satisfaction and self-esteem in men with ED treated with sildenafil |
| Coyne et al[22] | HIV-positive males who have sex with men | 486 | Validate an adapted version of IIEF-15 for use in HIV-positive men who have sex with men |
| Flynn et al[23] | Cancer patients | 389 | Validation of the PROMIS sexual function and satisfaction scales |
| García-Cruz et al[24] | Patients referred from general practitioners to urologic practice | 125 | Validate Erection Hardness Score in Spanish |
| Gelhorn et al[25] | Patients diagnosed with hypogonadism | 177 | Validate the Hypogonadism Impact of Symptoms Questionnaire Short Form |
| Gonzáles et al[26] | Patients participating in a cardiopulmonary or metabolic rehabilitation program | 78 | Validate the IIEF-15 in Portuguese (Brazil) in patients with cardiopulmonary and metabolic diseases |
| Hwang et al[27] | Males aged >30 | 1060 | Assess prevalence of erectile dysfunction in Taiwan |
| Kriston et al[28] | Patients with cardiovascular diseases in rehabilitation centers | 261 | Test 4 proposed factor structures of the IIEF-15 in German population |
| Maasoumi et al[29] | Males working in four different work settings | 181 | Validate the Sexual Quality of Life—Male in Persian (Iran) |
| Mulhall et al[30] | 190 men screened for ED ; 902 males participating in a community health survey | 1259 | Development of Sexual Experience Questionnaire |
| Nimbi et al[31] | Convenience sample | 425 | Validate the Sexual Modes Questionnaire in Italian |
| O'Toole[32] | Patients with inflammatory bowel disease | 175 | Develop a IBD-specific Male Sexual Dysfunction Scale |
| Parisot et al[33] | Patients with localized prostate cancer who underwent surgery | 75 | Validation and responsiveness of Erection Hardness Score |
| Pascoal et al[34] | Heterosexual males in a dyadic relationship | 129 | Development of the Beliefs About Sexual Functioning Scale |
| Quek et al[35] | 20 patients admitted for transurethral resection of the prostate and 20 control males | 40 | Validate the IIEF-15 in Malaysia |
| Quinta Gomes et al[36] | Sexually healthy males and patients with ED | 1363 | Validate the IIEF-15 in Portugal |

**Table 1.** Continued

| Reference | Population | Sample size | Main aim of study |
|---|---|---|---|
| Rosen et al[1] | 111 patients with ED part of a sildenafil RCT; 109 matched healthy men; 37 patients with ED; 21 matched healthy controls | 278 | Development and first validation of IIEF-15 |
| Rosen et al[37] | Participants in RCT on tadalafil | 863 | Estimate Minimal Clinically Important Difference for the Erectile Function subscale of the IIEF-15 |
| Rubio-Aurioles et al[38] | 51 couples with untreated ED; 57 couples without ED | 107 | Development and first validation of the Female Assessment of Male Erectile |
| Saffari et al[39] | Males attending a health post | 1764 | Validate the Male Genital Self-Image Scale for Iranian Men |
| Serefoglu et al[40] | Patients from an urology clinic | 430 | Analyze the impact of patient age, education level, and household income on the comprehension of the IIEF-15 (Turkish version) and determine the patient characteristics that make this questionnaire less reliable |
| Tang et al[41] | 260 patients diagnosed with premature ejaculation, and 104 healthy controls | 364 | Validate the Premature Ejaculation Diagnostic Tool in Chinese |
| Terrier et al[42] | Sexually active patients with early-stage prostate cancer after radical prostatectomy | 178 | Define the optimal Erectile Functioning score that optimally defines "functional" erections after radical prostatectomy |
| Wiltink et al[43] | 59 ED patients, 38 patients with Peyronie's disease, and 33 control males | 130 | Validate IIEF-15 for the German population (Germany) |
| **IIEF-15 & IIEF-5** | | | |
| Dargis et al[50] | Canadian males aged > 65 years | 508 | Validation of IIEF-15 and IIEF-5 in an older population |
| Lim et al[51] | 111 healthy males; 60 patients attending primary care clinics; 32 ED patients undergoing sildenafil therapy | 197 | Validate the IIEF-15 and IIEF-5 in Malay (Malaysia) |
| **IIEF-5** | | | |
| Aslan et al[44] | Patients with ED | 81 | Evaluate the association between IIEF-5 and Erection Hardness Grade Score in patients who underwent sildenafil citrate treatment for ED |
| Cappelleri et al[45] | Patients with ED enrolled in a RCT on sildenafil | 247 | Examine the relationship between patients' self-assessment of EF and classification of ED severity using the IIEF-5 |
| Lin et al[46] | Prostate cancer patients in sexual relationships | 1058 | Rasch analysis of Premature Ejaculation Diagnostic Tool and IIEF-5 in Iranian prostate cancer patients |
| Mahmood et al[47] | Patients from an urology clinic | 47 | Validate the IIEF-5 in Urdu (Pakistan) |
| Rosen et al[2] | 1063 patients with ED enrolled in a sildenafil RCT, and 116 healthy controls | 1152 | Development of an abridged version of the IIEF-15 (the IIEF-5) |
| Tang et al[48] | Patients diagnosed with LPE, heterosexual with a sexual relationship >6 months | 406 | Validate IIEF-5 for erectile function in Lifelong Premature Ejaculation patients in China |
| Utomo et al[49] | 82 ED patients; 253 controls | 335 | Validate IIEF-5 in Dutch (Netherlands) |

ED = erectile dysfunction; EF = erectile function; EHS = Erection Hardness Score; IBD = inflammatory bowel disease; IIEF = International Index of Erectile Function; LPE = lifelong premature ejaculation; PROMIS = Patient-Reported Outcomes Measurement Information System; QEQ = Quality of Erection Questionnaire; RCT = randomized controlled trial; SEAR = Self-Esteem And Relationship questionnaire; SEX-Q = Sexual Experience Questionnaire.

good",[1,16,22,28,36,38,50,51] "adequate",[26,31,43] or "inadequate".[16,34,35,41] The inadequate scores were due to only reporting internal consistency for the total IIEF-15 instead of its subscales[16,34,41] or because of a very small sample size ("other flaws" in COSMIN methodologic quality).[35]

8 studies, of "very good",[15,28,36,50] "adequate",[31] and "inadequate"[34,35,41] quality, reported Cronbach's $\alpha$ of sufficient values of the IIEF-15. 5 studies, of "very good"[1,22,38,51] and "adequate"[26] quality, reported Cronbach's $\alpha$ of insufficient values of the IIEF-15. In 2 studies, the evidence on internal

**Table 2.** Structural validity

| Reference | Methodology | Outcome | Rating | Quality |
|---|---|---|---|---|
| **IIEF-15** | | | | |
| Bushmakin et al[17] | Confirmatory factor analysis | 5-factor solution found on baseline (N = 500; CFI = .92); on end of DBPC phase (N = 458; CFI = .94); and end of open-label (N = 454; CFI = .93), all with bad fit (CFI < .95). | Insufficient | Very good |
| Coyne et al[22] | Principal component analysis | Four factors with Eigenvalue > 1.5. The original domains of intercourse and overall satisfaction appeared together in 1 factor. | Indeterminate | Adequate |
| Gonzáles et al[26] | Principal component analysis | 5 factors explaining 75.8% of variance; most questions were loaded correctly on their respective domains, except for sexual satisfaction domain, which comprises questions 6, 7, and 8, which presented a confounding factor. Question 1 equally loaded on 2 factors. | Indeterminate | Doubtful* |
| Kriston et al[28] | Confirmatory factor analysis | Original 5-factor model had acceptable fit (GFI = .889; TLI = .933; CFI = .949; SRMR = .045; RMSEA = .09) as did a 4-factor model (GFI = .849; TLI = .908; CFI = .926; SRMR = .049; RMSEA = .107). A 2-factor model had non-acceptable fit (CFI = .783; TLI = .854; CFI = .876; SRMR = .064; RMSEA = .134), as did a 1-factor model (GFI = .743; TLI = .812; CFI = .839; SRMR = .072; RMSEA = .152). CAIC favored the original 5-factor model (512.68). | Sufficient | Very good |
| Lim et al[51] | Principal component analysis | The expected structure of 5 distinct domains was not clearly present. The eigenvalue was concentrated on the first factor, whereas the remaining 4 factors extracted had eigenvalue <1. Factor 2 of the Malay version of IIEF corresponded with the OS domain of the original IIEF, whereas factor 3 corresponded with SD domain, and factor 4 with OF domain. Factor 1 contained a mixture of loadings from both EF and IS domains. | Indeterminate | Adequate |
| Quinta Gomes et al[36] | Principal component analysis | 2 components explaining 55% variance. The first component cluster loadings from 8 items of the erection and orgasm domains of the original IIEF. The second component included the original dimensions of SD, IS, and OS, was composed of the remaining 6 items of the scale. | Indeterminate | Doubtful† |
| Quinta Gomes et al[36] | Confirmatory factor analysis | Acceptable fit for 2-factor model (RMSEA = .077; CFI = .94; GFI = .93; AGFI = .90) and 5-factor model (RMSERA = .067; CFI = .96; GFI = .95; AGFI = .92) | Sufficient | Doubtful |
| Rosen et al[1] | Principal component analysis | Five factor solution. (1) erectile function, (2) orgasmic function, (3) sexual desire, (4) intercourse satisfaction, and (5) overall satisfaction. | Indeterminate | Adequate |
| Wiltink et al[43] | Principal component analysis | 2 factors found explaining 70% variance. First factor (12 items) of sexual function. Second factor (3 items) of sexual desire. | Indeterminate | Adequate |
| **IIEF-5** | | | | |
| Lin et al[46] | Rasch analysis | Monotonical increase across IIEF; 1 local dependency in IIEF; no substantial DIF in IIEF | Sufficient | Very good |

CFI = Comparative Fit Index; EF = erectile function; GFI = Goodness of Fit Index; IIEF = International Index of Erectile Function; IS = intercourse satisfaction; OF = orgasmic function; OS = overall satisfaction; RMSEA = Root Mean Square Error of Approximation; SD = sexual desire; SRMR = standardized root mean square residual; TLI = Tucker Lewis Index.
*Due to insufficient sample size.
†Due to very unequal subgroup sizes.

consistency was rated as indeterminate because it could not be interpreted: 1 study did not report the internal consistency per subscale,[16] and 1 study reported internal consistency for 2 subscales, resulting from their PCA results.[43]

5 studies reported on internal consistency of the IIEF-5[47–51] (Supplementary Table 1). Methodologic quality of these studies was rated as "very good"[48–51] or "inadequate".[47] The inadequate score was due to a very small number ("other flaws" in COSMIN methodologic quality).[47] The evidence of internal consistency was rated as indeterminate for all 5 studies, because unidimensionality was not investigated (see Structural Validity), which is a prerequisite for internal consistency.

## Test-Retest Reliability

8 studies reported on test-retest reliability of the IIEF-15[1,15,16,35,36,40,51] (Table 3). Methodologic quality of these studies was rated as "doubtful",[1,16,26,36,40,51] or "inadequate".[15,35] The doubtful scores were due to inappropriate time intervals (the same day)[40,51] and reporting of correlation coefficients instead of the intraclass correlation coefficient.[1,16,36,40] The inadequate scores were due to test conditions that differed across measurements,[15] and a very small number ("other flaws" in COSMIN methodologic quality).[35]

The evidence on test-retest reliability was rated as sufficient in 5 studies, of "doubtful"[1,26,51] and "inadequate"[15,35] quality. The evidence was rated as insufficient in 2 studies, of "doubtful"[36,40] quality, because reported values of reliability were <0.70. The evidence was rated as indeterminate in 1 study, of "doubtful"[16] quality, because the values were subdivided in 6 subgroups and not well interpretable.

2 studies reported on test-retest reliability of the IIEF-5.[49,51] Methodologic quality was rated as "adequate"[49] or "doubtful".[51] The doubtful score was due to inappropriate time intervals (the same day).[51] The evidence on test-retest reliability in both studies was rated as sufficient.

## Measurement Error

1 study reported measurement error of IIEF-15,[35] and measurement error was calculated for 1 study that reported test-retest reliability[1] (Supplementary Table 2). Methodologic quality was rated as "adequate"[1] or "inadequate".[35] The inadequate rating was due to a very small number ("other flaws" in COSMIN methodologic quality).[35]

For interpretation of measurement error, the minimal clinically important difference (MCID) is necessary. The evidence on measurement error was rated as indeterminate for the 2 studies[1,35] because no MCID was reported for any of the subscales in any of the included studies, except for the erectile function subscale for which a MCID was reported (mean MCID = 7.27).[37]

The evidence on measurement error of the erectile function subscale was rated as insufficient for 1 study,[35] for which we could calculate the standard error of measurement (0.69–3.59) and the smallest detectable change (SDC; 1.90–9.94). The SDC is the minimum change score necessary to have 95% confidence that it represents a true change. The MCID is the smallest change score that represents a clinically relevant change. The SDC should be smaller than the MCID, so that a smallest clinically relevant change score can be distinguished from measurement error. In this case, the SDC (9.49) was larger than the MCID (7.27), leading to an insufficient rating for the erectile function subscale.

1 study reported measurement error of the IIEF-5.[49] Methodologic quality was rated as "adequate." Limits of agreement (LoA) were reported (10.1). Evidence on measurement error was rated as indeterminate, because no MCID or MIC was reported.

## Construct Validity (Hypothesis Testing)

7 studies reported known-group comparison of the IIEF-15[1,35,36,41,43,50,51] (Supplementary Table 3). Known group differences were investigated in relation to age,[50] diagnosis of ED[1,36,43,51], diagnosis of premature ejaculation,[41] lifelong vs acquired premature ejaculation,[41] and treatment vs control.[35] The methodologic quality was rated as "adequate"[1,36,41,43,50,51] or "inadequate".[35] The inadequate rating was due to a very small number ("other flaws" in COSMIN methodologic quality).[35] Evidence for construct validity was rated as sufficient for all studies.

2 studies reported known-group comparison of the IIEF-5[2,50] and compared age groups[50] and diagnosis of ED.[2] The methodologic quality was rated as "adequate"[50] or "doubtful".[2] The doubtful rating was due to very unequal group sizes ("other flaws" in COSMIN methodologic quality).[2] Evidence of construct validity was rated as sufficient.

## Convergent Validity

17 studies reported on convergent validity of the IIEF-15[1,19,20,23–25,27,29–34,38,39,41,43] (Supplementary Table 4). The IIEF-15 was compared with a single-item self-assessment of ED,[19] the Patient Reported Outcomes Measurement Information System,[23] Quality Erection Questionnaire,[27] Erection Hardness Score,[20,24,27,33] Sexual Experience Questionnaire,[30] Male Genital Self-Image Scale,[39] Female Assessment of Male Erection,[38] partnership satisfaction,[43] Hypogonadism Impact of Symptoms Questionnaire Short Form,[25] Sexual Quality of Life—Male ,[29] Sexual Modes Questionnaire,[31] Inflammatory Bowel Disease Male Sexual Dysfunction Scale,[32] Beliefs About Sexual Functioning Scale,[34] Premature Ejaculation Tool,[41] and clinician ratings.[1,38,43]

The methodologic quality was rated as "adequate"[1,23,27,29,31,32,34,38,39,41,43] or "doubtful".[19,20,24,25,33] The doubtful ratings were due to a small number ("other flaws" in COSMIN methodologic quality),[33] use of the Pearson correlation where the Spearman correlation should have been

**Table 3.** Test-retest reliability

| Reference | Coefficient | IIEF-5 | Total score | EF | OF | SD | IS | OS | Rating | Quality |
|---|---|---|---|---|---|---|---|---|---|---|
| **IIEF-15** | | | | | | | | | | |
| Bayraktar et al[15] | Correlation | | .91 | .94 | .83 | .87 | .75 | .78 | Sufficient | Inadequate* |
| Bayraktar et al[16] | Rho | | .39–.87 | | | | | | Indeterminate | Doubtful† |
| Gonzáles et al[26] | ICC | | .80–.98 | .90–.98 | .91–.98 | .80–.92 | .82–.97 | .89–.98 | Sufficient | Doubtful |
| Quek et al[35] | ICC | | | .77 | .75 | .87 | .79 | .85 | Sufficient | Inadequate‡ |
| Quinta Gomes et al[36] | Correlation | | | .55 | .69 | .14 | .71 | .90 | Insufficient | Doubtful† |
| Rosen et al[1] | Correlation | | .82 | .84 | .64 | .71 | .81 | .77 | Sufficient | Doubtful† |
| Serefoglu et al[40] | Kappa | | .37 | | | | | | Insufficient | Doubtful* |
| **IIEF-15 & IIEF-5** | | | | | | | | | | |
| Lim et al[51] | ICC | .88 | .92 | .88 | .82 | .82 | .89 | .82 | Sufficient | Doubtful§ |
| **IIEF-5** | | | | | | | | | | |
| Utomo et al[49] | ICC | .88 | | | | | | | Sufficient | Adequate |

EF = Erectile Function; IIEF = International Index of Erectile Function; IS = intercourse satisfaction; OF = orgasmic function; OS = overall satisfaction; SD = sexual desire.

*Due to test conditions differing across measurements.

†Due to reporting of inappropriate coefficients.

‡Due to an extremely small number.

§Due to inappropriate time intervals.

used,[24] imprecise reporting of hypotheses ("other flaws" in COSMIN methodologic quality),[25] the lack of information on measurement properties of the comparator instrument,[19] or imprecise reporting of results.[20]

The evidence on construct validity was rated as sufficient for 11 studies, of "adequate"[1,23,27,29,30,38,43] and "doubtful"[19,24,25,33] quality. The evidence was rated as insufficient for 5 studies of "adequate"[31,32,34,39,41] and 1 study of "doubtful"[20] quality, because reported correlations were low.

2 studies reported on convergent validity of the IIEF-5,[44,45] and compared the IIEF-5 to the Erection Hardness Scale,[44] a single-item self-assessment of ED,[45] the Erectile Dysfunction Inventory of Treatment Satisfaction,[45] a 5-item version of the Erectile Dysfunction Inventory of Treatment Satisfaction filled in by a partner,[45] and a single item of global efficacy of erections.[45] Methodologic quality was rated as "adequate"[44] or "doubtful".[45] The doubtful rating was due to the lack of information on measurement properties of the comparator instrument.[45] The evidence on construct validity was rated as sufficient for 1 study[44] and insufficient for 1 study,[45] because the reported correlation was low.

## Divergent Validity

3 studies reported on divergent validity of the IIEF-15[1,43,50] (Supplementary Table 5) and compared the IIEF-15 to the Dyadic Adjustment Test and SF-12,[50] the Locke-Wallace Marital Adjustment Test,[1] State-Trait Anxiety Inventory, Center for Epidemiological Studies Depression Scale,[43] and social desirability.[1,43] Methodologic quality was rated as "adequate"[43,50] or "doubtful".[1] The doubtful score was due to non-reporting of measurement properties of the comparison instrument. The evidence on construct validity was rated as sufficient for all studies.

1 study reported on divergent validity of the IIEF-5[50] (Supplementary Table 5) and compared the IIEF-5 to the Dyadic Adjustment Test and SF-12. Methodologic quality was rated as "adequate," and evidence was rated as sufficient.

## Criterion Validity

4 studies reported on criterion validity of the IIEF-15 Erectile Function subscale[18,38,42,43] (Table 4). 1 study also reported criterion validity for the IIEF-15 total score.[43] Methodologic quality was "very good",[18,38] "adequate",[43] or "doubtful".[42] The "doubtful" rating was due to use of a questionable gold standard (intercourse satisfaction). All other studies used ED diagnosis as the gold standard.

The evidence on criterion validity was rated as sufficient for 3 studies of "very good"[18,38] and "doubtful"[42] quality. 2 studies[18,38] reported area under the curve (AUC) values for the erectile function subscale as 0.97 for diagnosing ED, with good sensitivity (0.97–0.98) and specificity (0.79–0.88) for the cut-off point of 25. 1 study[42] reported an AUC value for the erectile

**Table 4.** Criterion validity

| Reference | Instrument | AUC | Cutoff | Sensitivity | Specificity | PPV | NPV | Rating | Quality |
|---|---|---|---|---|---|---|---|---|---|
| **IIEF-15** | | | | | | | | | |
| Cappelleri et al[18] | IIEF-15 EF | .97 | 25 | .97 | .88 | .89 | .97 | Sufficient | Very good |
| Rubio-Aurioles et al[38] | IIEF-15 EF | .97 | 25 | .98 | .79 | | | Sufficient | Very good |
| Terrier et al[42] | IIEF-15 EF | .86 | 24 | .78 | .80 | | | Sufficient | Doubtful* |
| | | | 25 | .77 | .82 | | | | |
| Wiltink et al[43] | IIEF-15 Total | | 53 | .87 | .75 | .85 | | Indeterminate | Adequate |
| | IIEF-15 EF | | 21 | .84 | .72 | .84 | | | |
| **IIEF-5** | | | | | | | | | |
| Lim et al[51] | IIEF-5 | .86 | 17 | .85 | .75 | | | Sufficient | Adequate |
| Rosen et al[2] | IIEF-5 | .97 | 21 | .98 | .88 | .89 | .98 | Sufficient | Doubtful† |
| Tang et al[48] | IIEF-5 | .97 | 22 | 1.00 | .06 | | | Sufficient | Very good |
| | | | 15.5 | .97 | .86 | | | | |

AUC = area under the curve; CART = Classification and Regression Trees; IIEF = International Index of Erectile Function; NPV = negative predictive value; PPV = positive predictive value.

*Due to a doubtful criterion.

†Due to very unequal group sizes which biases the results of the CART algorithm; and due to usage of training sample in cross-validation.

function subscale as 0.86 for determining intercourse satisfaction. Good sensitivity (0.77 and 0.78) and specificity (0.92 and 0.80) were reported for the cutoff points of 24 and 25, respectively. The evidence was rated as indeterminate for 1 study,[43] because no AUC value was reported.

3 studies reported on criterion validity of the IIEF-5 [2,48,51] (Table 4). Methodologic quality was "very good",[48] "adequate",[51] or "doubtful".[2] The doubtful rating was due to very unequal group sizes.[2] The evidence on criterion validity was rated as sufficient for all studies, with reported AUC between 0.86−0.97.[2,48,51] All studies reported good sensitivity (0.85−0.98) and specificity (0.75−0.88) for cutoff points of 15.5, 17, and 21.

## Responsiveness

6 studies reported responsiveness of the IIEF-15[1,14,19,21,33,35] (Supplementary Table 6). Methodologic quality was rated as "adequate",[1,14,19,21,33] or "inadequate".[35] The inadequate rating was due to a very small number ("other flaws" in COSMIN methodologic quality).[35] The evidence on responsiveness was rated as sufficient for all 6 studies.

2 studies reported on responsiveness of the IIEF-5[45,49] (Supplementary Table 6). Methodologic quality was rated as "adequate"[45] or "doubtful".[49] The doubtful rating was due to a very small group of treated patients ("other flaws" in COSMIN methodologic quality). The evidence on responsiveness was rated as sufficient for both studies.

## Data Synthesis

The overall ratings of the measurement properties can be found in Table 5. Structural validity of the IIEF-15 was rated as inconsistent with evidence of moderate quality, due to the inconsistencies in the findings. Structural validity of the IIEF-5 was

rated as sufficient with evidence of moderate quality, because it was based on only 1 study.

Internal consistency of the IIEF-15 was rated as inconsistent with evidence of moderate quality because of inconsistencies in the findings. Internal consistency of the IIEF-5 was rated as indeterminate, because of the lack of evidence for unidimensionality.

Reliability of the IIEF-15 was rated as inconsistent with evidence of moderate quality, due to inconsistencies in the findings. Reliability of the IIEF-5 was rated as sufficient with evidence of moderate quality, due to some risk of bias resulting from the methodologic quality. For both IIEF-15 and IIEF-5, measurement error was rated indeterminate, except for the erectile function scale, which was rated as insufficient.

Construct validity (hypothesis testing) of the IIEF-15 was rated as inconsistent with evidence of moderate quality. 11 studies showed sufficient scores, whereas 6 studies showed insufficient scores. We note that some of the comparator instruments in convergent validity are of questionable relevance (eg, the Male Genital Self-Image Scale) or quality (eg, comparators that were only validated once in their lifetime). As such, while formally rating the construct validity of the IIEF-15 as inconsistent, the rating leans more to sufficient than insufficient. Construct validity of the IIEF-5 was rated as sufficient with evidence of high quality. 1 study showed values of insufficient convergent validity of the IIEF-5, these values were only just below sufficient levels and were discounted against the evidence for sufficient construct validity.

Criterion validity was rated as sufficient and evidence of high quality for the IIEF-15, and evidence of moderate quality for the IIEF-5 due to some risk of bias resulting from the methodologic quality. Responsiveness was rated as sufficient and evidence of high evidence for the IIEF-15 and as indeterminate for the IIEF-5.

**Table 5.** Ratings of measurement properties

| Measurement property | Rating of measurement property | Quality of evidence |
|---|---|---|
| **IIEF-15** | | |
| Structural validity | Inconsistent | Moderate |
| Internal consistency | Inconsistent | Moderate |
| Reliability | Inconsistent | Moderate |
| Measurement error | Indeterminate/Insufficient (Erectile Function subscale) | Very low |
| Construct validity | Inconsistent | Moderate |
| Criterion validity | Sufficient | High |
| Responsiveness | Sufficient | High |
| **IIEF-5** | | |
| Structural validity | Sufficient | Moderate |
| Internal consistency | Indeterminate | |
| Reliability | Sufficient | Moderate |
| Measurement error | Indeterminate | |
| Construct validity | Sufficient | High |
| Criterion validity | Sufficient | Moderate |
| Responsiveness | Indeterminate | |

## DISCUSSION

This systematic review investigated the evidence regarding the measurement properties of the IIEF-15 [1] and IIEF-5 [2]. In contrast to our hypothesis, most of the measurement properties were not rated as sufficient for both the IIEF-5 and IIEF-15. The IIEF-15 was rated as sufficient on criterion validity (of the Erectile Function subscale) and responsiveness, with sufficient ratings with high level of evidence. The evidence for structural validity, internal consistency, construct validity, and test-retest reliability were rated inconsistent, with moderate level of evidence. Measurement error for the Erectile Function subscale was rated as insufficient with very low quality of evidence, although it was indeterminate for the remaining subscales.

The IIEF-5 was rated as sufficient on criterion validity with high quality of evidence. The IIEF-5 was also rated as sufficient on structural validity, test-retest reliability, and construct validity, but with moderate quality of evidence because the evidence was based on very few studies. The evidence for internal consistency, measurement error, and responsiveness were rated as indeterminate.

With regard to structural validity, there is some evidence from CFAs[28,36] and PCAs[1,26] that the IIEF-15 consists of a 5-factor structure as hypothesized.[1] However, there is also evidence not supporting the 5-factor structure: 1 CFA found a poor fit for a 5-factor structure,[17] 1 CFA found acceptable fits for both a 2-factor (1 factor of erectile function and orgasm, and 1 factor of desire and satisfaction) and 5-factor structure,[36] 1 CFA found acceptable fits for both a 4-factor (combined factor of erectile

function and intercourse satisfaction) and a 5-factor structure,[28] and multiple PCAs found either a 4-factor solution (combined component of erectile function and intercourse satisfaction,[51] or combined component of intercourse satisfaction and overall satisfaction[22]), or a 2-factor solution (1 component of erectile function and orgasm, and 1 component of desire and satisfaction,[36] or 1 component of sexual function and 1 component of sexual desire[43]). There seems to be as much, if not more, evidence against the 5-factor structure.

The results of the current review are in line with the concerns raised by Forbes et al,[52,53] that the 5-factor structure is not as firmly established as argued by Rosen et al.[3,54] We agree with the reply by Rosen et al[54] that low correlations between subscales of the IIEF-15 do not warrant an insufficient rating of structural validity, but disagree with their underrating for the concerns regarding the structural validity of the IIEF-15. Their evidence cited concerns exploratory factor analyses, with no mention of confirmatory analyses that provide a higher level of evidence for structural validity. 2 of the confirmatory analyses we identified showed evidence for both the 5-factor structure and alternative factor structures,[28,36] and the remaining CFA showed evidence against the 5-factor structure.[17] Future studies are clearly needed to investigate alternative factor structures (eg, 2-factor, 4-factor, second-order hierarchical factors) and compare them directly to the posited 5-factor structure.

The structural validity of the IIEF-5 is also of interest. Whereas 1 Rasch analysis showed sufficient structural validity, no tests of unidimensionality were reported in any of the included articles. The IIEF-5 consists of items representing both erectile dysfunction (items 2, 4, 5, and 15 from the IIEF-15), as well as sexual intercourse satisfaction (item 7 from the IIEF-15). Theoretically, the IIEF-5 may be multidimensional due to the use of 2 constructs during development. Tests of unidimensionality are of importance to further determine the structural validity of the IIEF-5.

The internal consistency of the IIEF-15 showed values that were very high indicating possible redundancy ($\alpha > 0.95$; 3 studies of very good quality), as well as values considered too low ($\alpha < 0.70$; 1 study of very good quality). However, many studies (12 studies of inadequate to very good quality) showed sufficient internal consistency. The methodologic quality is of importance to put these values in context, where an equal number of very good–quality studies found insufficient as sufficient values. Considering these results, it is possible that internal consistency of the IIEF-15 may vary across subgroups. However, when examining the populations of the studies that reported sufficient values[16,28,31,34−36,41,50] vs those of the studies that reported insufficient values,[1,22,26,38,51] no clear pattern arose, with both groups of studies investigating different nationalities, as well as subgroups (eg, older men, HIV-positive men who have sex with men, sexually healthy men, men suffering from ED). Furthermore, these inconsistencies may be caused by differences in factor structure across subgroups. A future cross-cultural study design,

investigating measurement invariance, may help elucidate the inconsistencies of these findings.

The evidence on internal consistency of the IIEF-5 cannot yet be determined, because the unidimensionality (a prerequisite for internal consistency) has not yet been tested. However, if unidimensionality is tested and found to be sufficient, internal consistency is likely to be rated as sufficient. 1 study (of very good quality) found an insufficient value ($\alpha < 0.70$), whereas 3 studies of very good quality found sufficient values.

Although both the IIEF-15 Erectile Function subscale and the IIEF-5 were able to sufficiently predict ED diagnosis, it is not yet clear which cutoff scores are most suitable. Making a direct comparison between sensitivity and specificity ratings of cutoff scores across studies is beyond the score of the current review, because an individual patient meta-analysis would be required. Furthermore, a larger sample (ie, more studies investigating criterion validity) would be necessary for such a meta-analysis to provide a reliable result. Further investigation into the criterion validity of the IIEF-15 and IIEF-5 is necessary for a more nuanced interpretation.

More information is necessary regarding the measurement error of both the IIEF-15 and the IIEF-5. Currently, the only available evidence is based on 1 study of inadequate quality.[35] This evidence showed an insufficient value for the Erectile Function subscale, but it is not possible to determine whether this is an artifact of the poor methodology of the study. Given the high frequency of use of both the IIEF-15 (particularly the Erectile Function subscale) and the IIEF-5 in clinical screening for ED, as well as outcome measures for clinical trials, knowledge on measurement error is important to be able to determine whether clinical change (ie, clinical improvement or deterioration) is a true change or is an artifact of the measurement tool itself. Fortunately, 1 study of very good quality calculated the MCID using multiple methods on a very large sample.[37] This information can be used to interpret any measurement error that is calculated for the Erectile Function subscale. We recommend researchers performing a test-retest reliability designed to calculate the LoAs or SDC, to further inform the field. More studies investigating the MCID are also necessary to further interpret measurement error.

A limitation of this review is that we did not investigate content validity. Content validity needs to be established before other measurement properties can be regarded.[4] A future investigation of content validity is warranted. Another limitation of this review is the use of a precise rather than a sensitive search filter of measurement properties to identify studies to be included. The sensitivity of the precise filter was 93% in a random set of PubMed records, whereas the sensitivity of the sensitive search filter was 97%.[9] The use of the precise filter was a pragmatic choice over the available sensitive filter because the initial search encompassed 39 PROMs (including the IIEF-15 and IIEF-5), and the sensitive filter would provide too many hits for feasible screening. The possibility remains that the precise filter missed validation studies of the IIEF-15 and IIEF-5.

In 2002, the IIEF-15 was considered to "meet psychometric criteria for test reliability and validity".[3] We offer a more cautious interpretation of the measurement properties of the IIEF-15. Although we support the claim that the IIEF-15 meets psychometric criteria for criterion validity (in regard to the Erectile Function subscale) and responsiveness; we argue that structural validity, internal consistency, test-retest reliability, construct validity, and measurement error have not yet been demonstrated to meet psychometric criteria. Given the widespread of use of the IIEF-15 in both clinical practice and research, more thorough research is necessary regarding these measurement properties. A large-scale cross-cultural study design or an individual patient data meta-analysis, applying CFA, measurement invariance tests, internal consistency measures, and calculating the LoA or SDC, is recommended. It is possible that such research may suggest adjustments to be made to the IIEF-15 or its scoring.

The results of this review highlight a couple of important points for the interpretation of the IIEF-15 and IIEF-5 in clinical practice and research. First, some of the subscales may need to be combined, and interpreting them as 2 separate constructs may not be valid. Because the erectile function subscale is most often found in 1 factor with other subscales (based on both CFA and PCA), further research may find that other subscales should be combined with this subscale for a valid interpretation. Second, there is uncertainty what the optimal cutoff should be for the IIEF-15 and IIEF-5 to screen for ED, because multiple optimal cutoff scores were reported for both the IIEF-15 and IIEF-5. Further research is necessary to investigate optimal cutoff points. For current practice, it is important that researchers and clinicians maintain consistency, and, as such, the cutoff points of 25 for the IIEF-15 EF domain and 21 for the IIEF-5 should be maintained. We do suggest that researchers and clinicians keep a close eye on further research of criterion validity, because another cutoff point may prove to be more accurate. Third and last, the lack of information on measurement error is a problem for the interpretation of change scores of the IIEF-15 and IIEF-5. We advise using the IIEF in tandem with another measure when determining ED development in patients, because this may lead to a more robust interpretation of change over time.

## CONCLUSION

The IIEF-15 meets psychometric criteria for criterion validity (with regard to the Erectile Function subscale) and responsiveness; but structural validity, internal consistency, test-retest reliability, construct validity, and measurement error have not yet been demonstrated to meet psychometric criteria. In particular, further research into the structural validity of the IIEF-15 is of relevance. The IIEF-5 meets psychometric criteria for structural validity, test-retest reliability, construct validity, and criterion validity. Internal consistency, measurement error, and responsiveness require further research. The most pressing matter for future research is determining the unidimensionality of the IIEF-5.

## ACKNOWLEDGMENTS

**Corresponding Author:** Irma M. Verdonck-de Leeuw, Prof, Vrije Universiteit Amsterdam, Department of Clinical, Neuro- and Developmental Psychology, Amsterdam Public Health Research Institute, The Netherlands, Van Der Boechorststaat 7, 1081 BT Amsterdam, The Netherlands. Tel: +31 20 444 0931; Fax: +31 20 444 3688; E-mail: IM.Verdonck@vumc.nl

*Conflicts of Interest:* The authors declare no conflicts of interest.

## STATEMENT OF AUTHORSHIP

**Category 1**

**(a) Conception and Design**
Koen I. Neijenhuijs; Neil K. Aaronson; Bernhard Holzner; Caroline B. Terwee; Pim Cuijpers; Irma M. Verdonck-de Leeuw
**(b) Acquisition of Data**
Koen I. Neijenhuijs; Karen Holtmaat
**(c) Analysis and Interpretation of Data**
Koen I. Neijenhuijs; Karen Holtmaat

**Category 2**

**(a) Drafting the Article**
Koen I. Neijenhuijs
**(b) Revising It for Intellectual Content**
Koen I. Neijenhuijs; Karen Holtmaat; Neil K. Aaronson; Bernhard Holzner; Caroline B. Terwee; Pim Cuijpers; Irma M. Verdonck-de Leeuw

**Category 3**

**(a) Final Approval of the Completed Article**
Koen I. Neijenhuijs; Karen Holtmaat; Neil K. Aaronson; Bernhard Holzner; Caroline B. Terwee; Pim Cuijpers; Irma M. Verdonck-de Leeuw

## REFERENCES

1. Rosen RC, Riley A, Wagner G, et al. The International Index Of Erectile Function (IIEF): A multidimensional scale for assessment of erectile dysfunction. Urology 1997;49:822-830.
2. Rosen RC, Cappelleri JC, Smith MD, et al. Development and evaluation of an abridged, 5-item version of the International Index of Erectile Function (IIEF-5) as a diagnostic tool for erectile dysfunction. Int J Impot Res 1999;11:319-326.
3. Rosen R, Cappelleri J, Gendrano N Iii. The International Index of Erectile Function (IIEF): A state-of-the-science review. Int J Impot Res 2002;14:226-244.
4. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27:1147-1157.
5. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. Qual Life Res 2010;19:539-549.
6. Hout A, van der Uden-Kraan CF Van, Witte BI, et al. Efficacy, cost-utility, and reach of an eHealth self-management application "Oncokompas" that facilitates cancer survivors to obtain optimal supportive care: study protocol for a randomized controlled trial. Trials 2017;18:228.
7. Lubberding S, van Uden-Kraan CF Van, Te Velde EA, et al. Improving access to supportive cancer care through an eHealth application: A qualitative needs assessment among cancer survivors. J Clin Nurs 2015;24:1367-1379.
8. Jansen F, van Uden-Kraan CF, Van Zwieten V, et al. Cancer survivors' perceived need for supportive care and their attitude towards self-management and eHealth. Support Care Cancer 2015;23:1679-1688.
9. Duman-Lubberding S, van Uden-Kraan CF, Jansen F, et al. Feasibility of an eHealth application "OncoKompas" to improve personalized survivorship cancer care. Support Care Cancer 2016;24:2163-2171.
10. Terwee CB, Jansma EP, Riphagen II, et al. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009;18:1115-1123.
11. Mokkink LB, Terwee CB, Patrick DL, et al. COSMIN checklist manual. Amsterdam: VU University Medical Center; 2012.
12. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. Qual Life Res 2012;21:651-657.
13. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res May 2018;27:1171-1179.
14. Althof SE, O'Leary MP, Cappelleri JC, et al. Sildenafil citrate improves self-esteem, confidence, and relationships in men with erectile dysfunction: Results from an international, multicenter, double-blind, placebo-controlled trial. J Sex Med 2006;3:521-529.
15. Bayraktar Z, Atun AI. Despite some comprehension problems the international index of erectile function is a reliable questionnaire in erectile dysfunction. Urol Int 2012;88:170-176.
16. Bayraktar Z, Atun I. Impact of physician assistance on the reliability of the International Index of Erectile Function. Andrologia 2013;45:73-77.
17. Bushmakin AG, Cappelleri JC, Symonds T, et al. Further understanding of the International Index of Erectile Function at 15+ years: Confirmatory factor analysis and multidimensional scaling. Therap Innovation Regul Sci 2014;48:246-254.
18. Cappelleri J, Rosen R, Smith M, et al. Some developments on the international index of erectile function (IIEF). Drug Inform J 1999;33:179-190.
19. Cappelleri JC, Siegel RL, Osterloh IH, et al. Relationship between patient self-assessment of erectile function and the

erectile function domain of the International Index of Erectile Function. **Urology 2000;56:477-481**.

20. Cappelleri JC, Bushmakin AG, Symonds T, et al. Scoring Correspondence in Outcomes Related to Erectile Dysfunction Treatment on a 4-Point Scale (SCORE-4). **J Sex Med 2009; 9:809-819**.

21. O'Leary MP, Althof SE, Cappelleri JC, et al. Self-esteem, confidence, and relationship satisfaction in men with erectile dysfunction treated with sildenafil citrate: A multicenter, randomized, parallel-group, double-blind, placebo-controlled study in the United States. **J Urol 2006;175:1058-1062**.

22. Coyne K, Mandalia S, McCullough S, et al. The international index of erectile function: Development of an adapted tool for use in HIV-positive men who have sex with men. **J Sex Med 2010;7:769-774**.

23. Flynn KE, Reeve BB, Lin L, et al. Construct validity of the PROMIS sexual function and satisfaction measures in patients with cancer. **Health Qual Life Outcomes 2013;11:1**.

24. García-Cruz E, Romero Otero J, Martínez Salamanca JI, et al. Linguistic and psychometric validation of the erection hardness score to Spanish. **J Sex Med 2011;8:470-474**.

25. Gelhorn H, Roberts L, Khandelwal N, et al. Psychometric Evaluation of the Hypogonadism Impact of Symptoms Questionnaire Short Form (HIS-Q-SF). **J Sex Med 2017; 14:1046-1058**.

26. Gonzáles AI, Sties SW, Wittkopf PG, et al. Validation of the International Index of Erectile Function (IIFE) for use in Brazil. **Arquivos Brasil Cardiol 2013;101:176-181**.

27. Hwang TIS, Tsai TE, Lin YIC, et al. A survey of erectile dysfunction in Taiwan: Use of the erection hardness score and quality of erection questionnaire. **J Sex Med 2010;7:174**.

28. Kriston L, Günzler C, Harms A, et al. Confirmatory factor analysis of the German version of the International Index of Erectile Function (IIEF): A comparison of four models. **J Sex Med 2008;5:92-99**.

29. Maasoumi R, Mokarami H, Nazifi M, et al. Psychometric properties of the Persian translation of the Sexual Quality of Life-Male Questionnaire. **Am J Mens Health 2017;11:564-572**.

30. Mulhall JP, King R, Kirby M, et al. Evaluating the sexual experience in men: Validation of the sexual experience questionnaire. **J Sex Med 2008;5:365-376**.

31. Nimbi F, Tripodi F, Simonelli C, et al. Sexual Modes Questionnaire (SMQ): Translation and psychometric properties of the Italian version of the Automatic Thought Scale. **J Sex Med 2018;15:410-415**.

32. O'Toole A, de Silva PS, Marc LG, et al. Sexual dysfunction in men with inflammatory bowel disease: A new IBD-specific scale. **Inflamm Bowel Dis 2018;24:310-316**.

33. Parisot J, Yiou R, Salomon L, et al. Erection hardness score for the evaluation of erectile dysfunction: Further psychometric assessment in patients treated by intracavernous prostaglandins injections after radical prostatectomy. **J Sex Med 2014;11:2109-2118**.

34. Pascoal P, Alvarez M-J, Pereira C, et al. Development and initial validation of the beliefs about sexual functioning scale: A gender invariant measure. **J Sex Med 2017;14:613-623**.

35. Quek KF, Low WY, Razack AH, et al. Reliability and validity of the Malay version of the International Index of Erectile Function (IIEF-15) in the Malaysian population. **Int J Impot 2002; 14:310-315**.

36. Quinta Gomes AL, Nobre P. The International Index of Erectile Function (IIEF-15): Psychometric properties of the Portuguese version. **J Sex Med 2012;9:180-187**.

37. Rosen RC, Allen KR, Ni X, et al. Minimal clinically important differences (MCID) in the erectile function (EF) domain of the international index of erectile function (IIEF). **J Urol 2011; 185:e615**.

38. Rubio-Aurioles E, Sand M, Terrein-Roccatti N, et al. Female assessment of male erectile dysfunction detection scale (FAME): Development and validation. **J Sex Med 2009; 6:2255-2270**.

39. Saffari M, Pakpour AH, Burri A. Cross-cultural adaptation of the Male Genital Self-Image Scale in Iranian men. **Sex Med 2016;4:e34-e42**.

40. Serefoglu EC, Atmaca AF, Dogan B, et al. Problems in understanding the Turkish translation of the International Index of Erectile Function. **J Androl 2008;29:369-373**.

41. Tang D-D, Li C, Peng D-W, Zhang X-S. Validity of premature ejaculation diagnostic tool and its association with International Index of Erectile Function-15 in Chinese men with evidence-based-defined premature ejaculation. **Asian J Androl 2018;20:19-23**.

42. Terrier JE, Mulhall JP, Nelson CJ. Exploring the optimal erectile function domain score cutoff that defines sexual satisfaction after radical prostatectomy. **J Sex Med 2017;14:804-809**.

43. Wiltink J, Hauck EW, Phädayanon M, et al. Validation of the German version of the International Index of Erectile Function (IIEF) in patients with erectile dysfunction, Peyronie's disease and controls. **Int J Impot Res 2003;15:192-197**.

44. Aslan Y, Tuncel A, Aydin O, et al. The association between erection hardness grading scale and international index of erectile function in men with erectile dysfunction treated with sildenafil citrate. **Urol Int 2011;86:434-438**.

45. Cappelleri JC, Siegel RL, Glasser DB, et al. Relationship between patient self-assessment of erectile function and the Sexual Health Inventory for Men. **Clin Ther 2001;23:1707-1719**.

46. Lin CY, Pakpour AH, Burri A, et al. Rasch Analysis of the Premature Ejaculation Diagnostic Tool (PEDT) and the International Index of Erectile Function (IIEF) in an Iranian sample of prostate cancer patients. **PLoS ONE 2016;11:e0157460**.

47. Mahmood MA, Ur Rehman K, Khan MA, et al. Translation, cross-cultural adaptation, and psychometric validation of the 5-Item International Index of Erectile Function (IIEF-5) into Urdu. **J Sex Med 2012;9:1883-1886**.

48. Tang Y, Wang Y, Zhu H, et al. Bias in evaluating erectile function in lifelong premature ejaculation patients with the International Index of Erectile Function-5. **J Sex Med 2015; 12:2061-2069**.

49. Utomo E, Blok BF, Pastoor H, et al. The measurement properties of the five-item International Index of Erectile Function (IIEF-5): A Dutch validation study. **Andrology 2015;3:1154-1159.**

50. Dargis L, Trudel G, Cadieux J, et al. Validation of the International Index of Erectile Function (IIEF) and presentation of norms in older men. **Sexologies 2013;22:e20-e26.**

51. Lim TO, Das A, Rampal S, et al. Cross-cultural adaptation and validation of the English version of the International Index of Erectile Function (IIEF) for use in Malaysia. **Int J Impot Res 2003;15:329-336.**

52. Forbes MK, Baillie AJ, Schniering CA. Critical flaws in the Female Sexual Function Index and the International Index of Erectile Function. **J Sex Res 2014;51:485-491.**

53. Forbes MK. Response to Rosen et al. (2014) "Commentary on 'Critical Flaws in the FSFI and IIEF." **J Sex Res 2014; 51:498-502.**

54. Rosen RC, Revicki DA, Sand M. Commentary on "Critical Flaws in the FSFI and IIEF". **J Sex Res 2014; 51:492-497.**

## SUPPLEMENTARY DATA

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jsxm.2019.04.010.